Environmental Health Perspectives
Vol. 32, pp. 157-167, 1979

Statistical Analyses of the Relative Risk

by John J. Gart*

Let P_1 be the probability of a disease in one population and P_2 be the probability of a disease in a second population. The ratio of these quantities, $R = P_1/P_2$, is termed the relative risk.

We consider first the analyses of the relative risk from retrospective studies. The relation between the relative risk and the odds ratio (or cross-product ratio) is developed. The odds ratio can be considered a parameter of an exponential model possessing sufficient statistics. This permits the development of exact significance tests and confidence intervals in the conditional space. Unconditional tests and intervals are also considered briefly. The consequences of misclassification errors and ignoring matching or stratifying are also considered. The various methods are extended to combination of results over the strata. Examples of case-control studies testing the association between HL-A frequencies and cancer illustrate the techniques.

The parallel analyses of prospective studies are given. If P_1 and P_2 are small with large sample sizes the appropriate model is a Poisson distribution. This yields a exponential model with sufficient statistics. Exact conditional tests and confidence intervals can then be developed. Here we consider the case where two populations are compared adjusting for sex differences as well as for the strata (or covariate) differences such as age. The methods are applied to two examples: (1) testing in the two sexes the ratio of relative risks of skin cancer in people living in different latitudes, and (2) testing over time the ratio of the relative risks of cancer in two cities, one of which fluoridated its drinking water and one which did not.

Retrospective Studies

McKinlay (*1*) has reviewed the more general aspects of the design and analysis of retrospective studies.

The Odds Ratio

Let p_1 be the proportion of individuals of one population (smokers, genetic trait carriers, etc.) and q_1 be the proportion of the second population (e.g. nonsmokers, etc.) among the "cases." Similarly define p_2 and q_2 in the controls. The ratio or cross-product ratio was defined by Fisher (*2*) to be:

$$\psi = (p_1 q_2)/(p_2 q_1) \tag{1}$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. Let π be the frequency of the disease (or cause of death) in the total of the two populations.

Consider the relation of ψ to R. Using Bayes' theorem, we have, following Cornfield (*3,4*) and arguments implicit in Doll and Hill (*5*),

$$R = \frac{P_1}{P_2} = \frac{(p_1\pi)/[p_1\pi + p_2(1-\pi)]}{(1-p_1)\pi/[(1-p_1)\pi + (1-p_2)(1-\pi)]} \tag{2}$$

$$R = \frac{p_1}{1-p_1}\left[\frac{(1-p_2) + (1-p_1)\pi/(1-\pi)}{p_2 + p_1\,\pi/(1-\pi)}\right] \tag{3}$$

If the disease or cause of death is rare, $\pi/(1-\pi) \simeq 0$ and $R \simeq \psi$. Sometimes the relative risk and the odds ratio are used as interchangeable terms. It should be noted that this argument depends on the rare disease assumption (*6-8*). The important point is that ψ can be calculated from p_1 and p_2, which can be estimated from a case-control study.

The odds ratio has a history of use in statistical theory independent of this argument. Fisher (*2*) showed that ψ is the parameter in the noncentral distribution of the exact conditional test. The definition of interaction in $2 \times 2 \times k$ contingency tables (*9,10*) is the equality of the odds ratio over the several 2×2 tables. Other interesting properties of this measure are given in Armitage's (*11*) review.

*Mathematical Statistics and Applied Mathematics Section, National Cancer Institute, Bethesda, Md. 20014 U.S.A.

Conditional Inference on ψ.

Let x_1 be the number of individuals of the first population in a sample of n_1 cases and let x_2 be the number of individuals of the first population in a sample of n_2 controls. We assume x_1 and x_2 to be independent binomial variates with parameters p_1 and p_2, respectively. It is convenient to reparameterize to a logistic form,

$$p_1 = \frac{\exp\{\beta + (\lambda/2)\}}{1 + \exp\{\beta + (\lambda/2)\}}$$
$$p_2 = \frac{\exp\{\beta - (\lambda/2)\}}{1 + \exp\{\beta - (\lambda/2)\}} \quad (4)$$

This implies that $\psi = e^{\lambda}$ and that $\lambda = \ln(p_1/q_1) - \ln(p_2/q_2)$, the difference of the logits. In other words, inferences on λ are equivalent to those on ψ (or R under the rare disease assumption). The likelihood for this model is:

$$L(\lambda,\beta) = \binom{n_1}{x_1}\binom{n_2}{x_2}\exp\{[(x_1 - x_2)\lambda/2] + (x_1 + x_2)\beta\}g(\lambda,\beta) \quad (5)$$

The form of this likelihood implies that x_1 and $x_1 + x_2$ are a minimal set of sufficient statistics. Inferences on λ should be made by considering the conditional likelihood of x_1 given $x_1 + x_2 = x.$ is fixed (*12*).

This yields

$$h(x_1|x.;\psi) = \frac{\binom{n_1}{x_1}\binom{n_2}{x. - x_1}\psi^{x_1}}{\sum_{i=0}^{x.}\binom{n_1}{i}\binom{n_2}{x. - i}\psi^{i}} \quad (6)$$

as first noted by Fisher (*2*). One may use Eq. (6) to test exactly a hypothetical value of ψ, say $H_0 : \psi = \psi_0$ vs. $H_1 : \psi_1 > \psi_0$. The P value is then

$$P = \sum_{i=x_1}^{x.} h(i|x.;\psi_0) \quad (7)$$

If $\psi_0 = 1(\lambda = 0$ or $p_1 = p_2)$, $h(x_1|x.;1)$ reduces to the hypergeometric and the Fisher-Irwin exact test results. Exact confidence intervals (with confidence coefficients $\geq 1 - \alpha$) are calculated from Eqs. (8)

$$\sum_{i=0}^{x_1} h(i|x.;\psi_U) = \alpha/2$$
$$\sum_{i=x_1}^{x.} h(i|x.;\psi_L) = \alpha/2 \quad (8)$$

where ψ_U and ψ_L are the upper and lower limits, respectively (*4,13*). Note that these limits will always agree with the exact significance test in the sense that if $P < \alpha/2$, $\psi_0 \notin (\psi_L, \psi_U)$ and if $P > \alpha/2$, $\psi_0 \in (\psi_L,\psi_U)$. Cox (*14*) [see also Birch (*15*)] showed that the conditional maximum likelihood estimator of ψ, $\hat{\psi}_{cml}$, is given by the solution to the equation,

$$x_1 = E(X_1|x.;\hat{\psi}_{cml}) \quad (9)$$

where X_1 indicates the random variable whose expectation is calculated from Eq. (6). It is interesting to note that this is not, in general, equal to $\hat{\psi}_{aml} = \hat{p}_1\hat{q}_2/(\hat{p}_2\hat{q}_1)$, where the usual sample proportions are substituted for these parameters. Computer programs for calculating all these exact conditional methods are available (*16,17*).

The asymptotic approximations to these methods are rather easy to apply. The chi-square test of $\psi = 1$ is calculated by using the first two moments of the hypergeometric distribution. With this continuity correction, this is

$$X^2_{corr.} = \frac{[|x_1 - E(X_1|x.;\psi = 1)| - \tfrac{1}{2}]^2}{V(X_1|x.;\psi = 1)}. \quad (10)$$

For the general case of $\psi = \psi_0$, it is necessary to find the value of $\hat{x}_1$, for which, in the asymptotic distribution

$$E(X_1|x.;\psi_0) = \hat{x}_1 \quad (11)$$

This is the solution to the equation

$$\psi_0 = \frac{\hat{x}_1(n_2 - x. + \hat{x}_1)}{(n_1 - \hat{x}_1)(x. - \hat{x}_1)} \quad (12)$$

Here we assume, implicitly, that $x. \leq n_1$, and that we choose the solution to the quadratic, $\hat{x}_1$, such that $0 \leq \hat{x}_1 \leq \min(n_1,x.)$. The variance of X_1 is given by (*4,18*),

$$\hat{V}(X_1 \mid x.;\psi_0) = \left(\frac{1}{\hat{x}_1} + \frac{1}{x. - \hat{x}_1} + \frac{1}{n_1 - \hat{x}_1} + \frac{1}{n_2 - x. + \hat{x}_1}\right)^{-1} \quad (13)$$

An approximate chi-square test of $\psi. = \psi_0$ is given by

$$\chi^2_{corr}(\psi_0) = \frac{(|x_1 - \hat{x}_1| - \tfrac{1}{2})^2}{\hat{V}(X_1|x.;\psi_0)}. \quad (14)$$

The asymptotic maximum likelihood estimator is

$$\hat{\psi}_{aml} = x_1(n_2 - x_2)/[x_2(n_1 - x_1)] \quad (15)$$

The asymtotic confidence limits are found by solving the quartic equations (*4,13*),

$$\chi_\alpha^2 = \frac{\{x_1 - \hat{E}(X_1|x.;\psi) \pm \frac{1}{2}\}^2}{\hat{V}(X_1|x.;\psi)}, \quad (16)$$

where χ_α^2 is the upper percentage point of a chi-square variate with one degree of freedom. The smallest root with the $-\frac{1}{2}$ is denoted by $\hat{x}_{1L}$ and the largest with the plus sign is denoted by $\hat{x}_{1U}$. The approximate 1-α confidence limits are then

$$\hat{\psi}_L = \frac{\hat{x}_{1L}(n_2 - x. + \hat{x}_{1L})}{(x. - \hat{x}_{1L})(n_1 - \hat{x}_{1L})},$$

and

$$\psi_U = \frac{\hat{x}_{1U}(n_2 - x. + \hat{x}_{1U})}{(x. - \hat{x}_{1U})(n_1 - \hat{x}_{1U})}. \quad (17)$$

As an alternative to solving quartics, Gart (*19*), following a suggestion by Cornfield (*4*), has given a much simpler method for solving Eq. (16) based on an iterative procedure involving only the solution of quadratics. Gart and Thomas (*20*) assessed, in the conditional space, the various approximate methods of finding confidence limits for ψ. They found Cornfield's method [Eq. (14)] to be clearly superior to logit or Fieller Theorem methods. They recommended that Cornfield's method be used for 95% limits if the minimum of $\hat{x}_1, x. - \hat{x}_1, n_1 - \hat{x}_1$, and $n_2 - x. + \hat{x}_1$ in Eqs. (17) exceeds 1. For 99% limits this minimum should exceed 3. Otherwise an exact calculation, an approximate graphical solution (*21*) or Gart's (*22*) Method II is recommended.

Unconditional Inference on ψ

If we wish to consider inferences in the unconditional space (where $x.$ is not considered fixed) the asymptotic methods considered above still hold. It is appropriate to drop the ½ correction from the X^2_{corr} and from the calculation of the confidence coefficients [Eq. (16)]. As before, the uncorrected chi-square tests will also agree with uncorrected confidence limits in having $P < \alpha/2$ when the limits do not cover ψ_0 and visa-versa.

Recently Miettinen (*6*) has introduced a "test-based" method for finding confidence limits. For the relative risk this involves first calculating the chi-square test of the 2 x 2 table, X^2. The confidence limits for ψ are similar to those based on a logit transformation with the logit variances estimated indirectly from the X^2 test statistic:

$$\psi_{U,L} = \hat{\psi}_{aml}^{\,1 \pm (\chi_{\alpha/2}/X)} \quad (18)$$

This method will always agree in covering $\psi = 1$, when $\hat{P} > \alpha/2$ and not covering $\psi = 1$ when $\hat{P} < \alpha/2$, where $\hat{P}$ is calculated from the chi-square test. It is not clear from Miettinen's papers (*6,7*) whether he is recommending his method in the conditional or unconditional space. Since it is essentially the logit method, it is likely to perform similarly to that method in the conditional space. Gart and Thomas (*20*) found the logit method to yield much too narrow limits (i.e., true confidence coefficient < 1-α) in the conditional space. In the unconditional space Halperin (*8*) pointed out that Miettinen's argument strictly holds only when $\psi = 1$ (or $\psi = \psi_0$ if the chi-square is applied to testing $\psi = \psi_0$). He stated that asymptotically the true confidence coefficient will be less than 1-α for any $\psi \neq 1$ (or ψ_0). Miettinen (*7*) replied "that the *principle*, properly understood, is flawless . . .", and that a "systematic evaluation of the accuracy of test-based confidence limits are [sic] obviously needed." Apparently he claims his limits are adequate for $\psi \sim 1$. Gart and Thomas (*23*) have evaluated exactly (by enumeration of all the points) the lower limit of the unconditional confidence coefficient for both Cornfield's method [Eq. (16)] (omitting the ½ correction) and Miettinen's test based limits, [Eq. (18)] omitting the continuity correction from the chi-square test). The results are given in Table 1.

It is seen that the Cornfield's and Miettinen's method have the same confidence coefficients for $p_1 = p_2$ or $\psi = 1$. However, as p_1 and p_2 diverge, Cornfield's method is as close or closer to 0.95 than is Miettinen's method. In every case of Table 1 where $\psi \neq 1$, Miettinen's method yields coefficients less than 0.95 and the deviation is quite large for $\psi >> 1$. In such cases we see the test-based limits are not correct in principle, in asymptotic theory, nor in exact evaluation. One can only agree with Halperin's (*8*) assertion that they "should not be used since methods not suffering this disability, even if computationally somewhat more onerous, are always available." The exact or Cornfield's limits can be quickly found by program (*16,17*), and the iterative method suggested by Gart (*19*) is not that difficult to do on a small pocket calculator. In any case, the true appropriate "test-based" interval is the modification of the chi-square test given by Cornfield's method [Eq. (16)], since that equation reduces to the usual chi-square statistics [Eq. (10)] when $\psi = 1$. Fisher (*13*) showed this relation very clearly.

Misclassification Errors

All the results given above assume only sampling variation is present. The cases and controls are assumed to be classified correctly as to their population (smoking status, genetic type, etc.). Consider now the effect of misclassification. In the cases let θ_1

Table 1. Exact upper bounds of the confidence coefficients of Cornfield's (C) and Miettinen's (M) 95% limits on ψ in the unconditional sample space for samples of size 50 ($n_1 = n_2 = 50$)[a]

p_2	p_1	C	M
0.01	0.01	0.997	0.997
	0.05	0.942	0.752
	0.10	0.946	0.505
	0.25	0.947	0.333
	0.50	0.944	0.323
	0.75	0.950	0.321
	0.90	0.960	0.320
0.05	0.05	0.956	0.956
	0.10	0.962	0.814
	0.15	0.950	0.867
	0.25	0.960	0.885
	0.50	0.961	0.865
	0.75	0.961	0.842
	0.90	0.960	0.813
0.10	0.10	0.949	0.949
	0.15	0.947	0.858
	0.20	0.951	0.902
	0.25	0.950	0.932
	0.50	0.950	0.915
	0.75	0.949	0.901
	0.90	0.942	0.883
0.25	0.25	0.949	0.949
	0.50	0.952	0.945
	0.75	0.951	0.933
	0.90	0.949	0.943
0.50	0.50	0.943	0.943
	0.75	0.952	0.945
	0.90	0.950	0.915

[a]Data of Gart and Thomas (*23*).

be the probability that a first population individual (e.g., smoker) be classified in the second population (e.g., nonsmoker), and let ϕ_1 be the probability that a second population individual (e.g., nonsmoker) be classified as a first population individual (e.g., a smoker). In the controls let the corresponding probabilities be θ_2 and ϕ_2. The apparent odds ratio, ψ_a, would then be given

$$\psi_a = \frac{(1-\theta_1)p_1 + \phi_1 g_1}{(1-\theta_2)p_2 + \phi_2 q_2}\left[\frac{1-(1-\theta_2)p_2 - \phi_2 q_2}{1-(1-\theta_1)p_1 - \phi_1 q_1}\right] \tag{19}$$

Bross (*24*), in a significant paper, considered the case where the misclassification errors were equal in the cases and controls: $\theta_1 = \theta_2 = \theta$ and $\phi_1 = \phi_2 = \phi$.

It can then be shown that

$$\psi_a = \left[\frac{p_1}{p_2} - \frac{q_1\phi(\psi-1)}{p_2(1-\theta)+\phi q_2}\right]\left[\frac{q_2}{q_1} - \frac{p_2\,\theta(\psi-1)}{q_1(1-\phi)+\theta p_2}\right] \tag{20}$$

If we define the populations such that $\psi > 1$, it immediately follows the $\psi_a \leq \psi$, the true odds ratio. The equality holds only when $\psi = 1$ or $\phi = \theta = 0$. Thus under this model the observed odds ratio can only underestimate the true odds ratio.

If $\theta_1 \neq \theta_2$ and or $\phi_1 \neq \phi_2$ it is not possible to make any general statements. Depending on the parameters, ψ_a may be less than or greater than ψ. Goldberg's (*25*) paper is an excellent discussion of the problems of misclassification and contains an extensive bibliography [see also Copeland et al. (*26*)].

Heterogeneity of Response

The methods derived above assume a binomial distribution. This implies that in both the case and control groups each individual has the same probabilities, p_1 and p_2, respectively of being from population 1. In practice this is seldom the situation. Quite often, the cases and controls are matched or paired by age, sex, and other factors. Then it is reasonable to consider not a single p_1 and a single p_2 but a series of them for each pair, p_{1k}, $k = 1, 2, \ldots, n$, and p_{2k}, $k = 1, 2, \ldots, n$. The usual model assumed is that $\psi = (p_{1k}q_{2k}) / (p_{2k}/q_{1k})$, i.e., that the odds ratio is constant over the pairs (*27*). The pairing is often ignored in estimating the odds ratio; the results being pooled into a single 2 × 2 table. Let $\hat{\psi}_p$ be the pooled estimator. Armitage (*11*) showed that asymptotically

$$E(\hat{\psi}_p) = \psi\, \bar{w}/\bar{w}' \tag{21}$$

where $w_k = p_{2k}/q_{2k}$, $\bar{w} = (\Sigma q_{1k}w_k)/(\Sigma q_{1k})$, and $\bar{w}' = (\Sigma q_{2k}w_k)/(\Sigma q_{2k})$. Using the fact that $q_{2k}/q_{1k} = (1 + w_k\psi)/(1 + w_k)$ is an increasing function of w_k, Armitage goes on to show that $\bar{w}' > \bar{w}$ and thus that $1 < E(\hat{\psi}_p) < \psi$. Siegel and Greenhouse (*28*) showed a slightly less general result and give other references on the topic. Thus it is concluded that pooling can only underestimate the odds ratio.

Differential Bias in the Section of Cases and Controls

In the case we have just considered the heterogeneity of the response was balanced by having a one-to-one correspondence in the matched pairs. If there are differing biases in the selection of cases and controls, the direction of the bias in the pooled odds ratio cannot be predicted. Consider the rather extreme, hypothetical examples (29) shown in Table 2. It is seen that the pooled estimator may give a completely different picture than the estimators from the individual strata. We consider appropriate models for the combination of strata in a subsequent section.

Table 2

k	Example 1			Example 2		
	x_{1k}/n_{1k}	x_{2k}/n_{2k}	$\hat{\psi}_k$	x_{1k}/n_{2k}	x_{2k}/n_{2k}	$\hat{\psi}_k$
1	10/20	150/300	1.00	15/20	160/300	2.63
2	250/300	50/60	1.00	150/300	5/20	3.00
Totals	260/320	200/360	$\hat{\psi}_p = 3.47$	165/320	165/320	$\hat{\psi}_p = 1.00$

Application

The human leukocyte antigen system (HL-A) is a genetic system having several alleles. The allele HL-A2 was implicated by several authors as predisposing children to acute lymphocytic leukemia (ALL). Rogentine et al. (*30*) did a case-control study with fifty patients (n_1) who were examined at the National Cancer Institute. The control group were 200 (n_2) normal blood donors. As the genetic trait is autosomal, there was no need for matching on sex. Race was considered; all the cases and controls are white. Among the cases, 36 (x_1) individuals belonged to the HL-A2 population (72%) and among the controls 83 (x_2) possessed this antigen (41.5%). For these numbers we find from Eq. (9) that $\hat{\psi}_{\text{cml}} = 3.61$ and $\hat{\psi}_{\text{aml}} = [(36)\ (117)]/[(83)\ (14)] = 3.62$. The exact test of $\psi = 1$ yields a one-tailed $P = 0.00009$ while the approximate test [Eq. (9)] yields $P = 0.00011$. The 95% and 99% confidence limits are given in Table 3. It is noted that both the exact and approximate P are < 0.005, and thus both corresponding 99% confidence intervals exclude $\psi = 1$. Note also that the discrepancies between the exact and approximate methods are small throughout.

One explanation for the observed association between HL-A2 and ALL considered by Rogentine et al. (*30*) was that possession of HL-A2 conferred susceptibility to ALL. However, the 50 patients typed for the HL-A system were not an unbiased sample of the patients admitted over the period of years considered. In a later paper Rogentine et al. (*31*) analyzed a larger series of 137 patients cared for at NIH between 1962 and 1971. The HL-A typing was not begun until 1969 so that those diagnosed between 1962 and 1968 must necessarily have survived on year. For 1962-69, there were 32 in the typed series among the total of 279 patients (or 11%). In the 1969-71, the 53 typed cases represent 52% of the cases in that time period. If we look at the HL-A2 frequency by year of diagnosis we find the results shown in Table 4. There is no significant difference between the 1969-71 frequency and the controls while the 1962-68 frequency is significantly different from both the 1969-71 frequency ($P < 0.01$) and the control frequency ($P < 0.001$). Furthermore, it was found that of the eighteen patients surviving 1500 days from diagnosis 15 (94 83%) had HL-A2. It was concluded that HL-A2 may not confer susceptibility to HL-A2, but, in fact, may be associated with longer survivorship.

Whenever one selects by a nonrandom mechanism from a larger set of cases a danger of bias is present. Here there is a bias in calendar date of admission and the fact that the cases must survive to the time the study began. Also they must survive to reach a research hospital.

Table 3

	95%	99%
Exact:	$1.77 \le \psi \le 7.72$,	$1.45 \le \psi \le 9.79$,
Approximate:	$1.76 \le \psi \le 7.57$,	$1.43 \le \psi \le 9.28$.

Combination over Strata

Consider now the situation where the cases and controls may each be divided into several matched strata or blocks. Let $k = 1, 2, \ldots, K$ be the index for the strata. Extend the notation used above by adding a subscript k, i.e., $p_{1k}, p_{2k}, n_{1k}, n_{2k}, \beta_k; k = 1, 2, \ldots, K$. However, let the λ remain constant over the strata. This implies the $\psi_k = (p_{1k}q_{2k})/(p_{2k}/q_{1k}) = e^{\lambda} = \psi$, is constant over the strata. Thus we are interested in making inferences on the common odds ratio, ψ, in the several 2×2 tables. Other models as possible, but this one has several convenient consequences: (1) it leads to a minimal set of sufficient statistics for which conditional inference is valid; (2) it is equivalent to the "no interaction" model of Bartlett (*9*) for

Table 4. HL-A2 frequency among all cases by year of diagnosis.

Year of diagnosis	Proportion with HL-A2 (%)
1962-68	27/32 (84%)
1969-71	28/53 (53%)
Controls	177/401 (44%)

$2 \times 2 \times K$ contingency tables; (3) Radhakrishna (*32*) has shown this model to be robust with respect to asymptotic efficiency relative to other possible models, such a difference in logarithms or the arc sins of the square root of the p; (4) Asymptotic unconditional inference can be conveniently employed using the logit transformation (*33*).

The methods for analyzing this model has been extensively reviewed by Gart (*19*). We shall limit ourselves here to sketching the main points.

The extended model is easily shown to possess a minimal set of $K+1$ sufficient statistics: $x_{1.} = \Sigma_k x_{1k}$ and $x_{.k}$, $k = 1, 2, \ldots, K$. Optimal inferences on the common ψ (or λ) can then be based on the conditional distribution of $x_{1.}$ given the $x_{.k}$ are all fixed. This leads to a product of K independent generalized hypergeometric distributions [Eq. (6)]:

$$h(x_{11}, x_{12}, \ldots, x_{1k}|x_{.1}, x_{.2}, \ldots, x_{.k}; \psi) = \Pi_k h_k(x_{1k}|x_{.k}; \psi) \quad (22)$$

Exact tests of $\psi = \psi_0$ (or $\psi = 1$) and exact confidence limits for ψ are found from Eq. (22) in much the same manner as for the single stratum case (*34,35*). The conditional maximum likelihood estimator (*15*) is the solution to Eq. (23):

$$x_{1.} = E(X_{1.}|x_{.1}, x_{.2}, \ldots, x_{.K}; \hat{\psi}_{\text{cml}}) \quad (23)$$

in parallel with the previous result. In addition here, an exact test of the model is that ψ is constant. This test is based on the conditional distribution of x_{11}, $x_{12}, \ldots x_{1K}$, given $x_{1.}$ and $x_{.1}, x_{.2}, \ldots x_{.2}, \ldots x_{.K}$ are all fixed (*9*). All these conditional "exact" methods are implemented in a program by Thomas (*17*).

It shall be noted that the above theory yields the usual methodology for the matched case, $n_{1k} = n_{2k} = 1$ for all k. An exact form of McNemar's test is given by Eq. (22), and $\hat{\psi}_{\text{cml}}$ is the ratio of unlike pairs (*27*).

The asymptotic methodology can be extended rather easily. Cochran (*34*) introduced the approximate test for $\psi = 1$, based on a chi-square statistic. We give here the somewhat modified form [due to Mantel and Haenszel (*36*)] of Cochran's test:

$$X^2_{\text{corr}} = \frac{(|x_{1.} - \Sigma_k E(X_{1k}|x_{.k}; \psi = 1)| - ½)^2}{\Sigma_k V(X_{1k}|x_{.k}; \psi = 1)} \quad (24)$$

where the means and variances are the computed from the hypergeometric distribution.

Gart (*37*) discussed the asymptotic noncentral conditional distribution. For the general case of $\psi = \psi_0$, we need to find the values of $\hat{x}_{1k}$, $k = 1, 2, \ldots, K$, for which, in the asymptotic distribution, we have

$$E(X_{1k}|x_{.k}; \psi_0) = \hat{x}_{1k} \qquad k = 1, 2, \ldots, K$$

This involves solving K quadratic equations equivalent to Eq. (12). The mean of the asymptotic normal distribution of $X_{1.}$ is then,

$$\hat{E}(X_{1.}|x_{.1}, \ldots x_{.k}; \psi_0) = \Sigma_k \hat{x}_{1k} \quad (25)$$

and the variance is

$$\hat{V}(X_{1.}|x_{.1}, \ldots x_{.k}; \psi_0) = \Sigma_k \hat{V}(X_{1k}|x_{.k}; \psi_0) \quad (26)$$

where the $\hat{V}(X_{1k} \mid x_{.k}; \psi_0)$ is given by an equation similar to Eq. (13). The test for $\psi = \psi_0$ is found from a formula equivalent to Eq. (14) and confidence limits from formulas equivalent to Eqs. (16) and (17) [(see also Gart (37)]. The asymptotic maximum likelihood estimator is given by the solution (*37*) to Eq. (27)

$$x_{1.} = \hat{E}(X_{1.}|x_{.1}, x_{.2}, \ldots x_{.k}; \hat{\psi}_{\text{aml}}) \quad (27)$$

which involves the simultaneous solution of K quadratic equations.

The set of equations solved in finding $\hat{\psi}_{\text{aml}}$ is the same as those cited by Bartlett (*9*) and Norton (*10*) in deriving the approximate chi-square test for interaction in $2 \times 2 \times K$ contingency tables. The usual goodness of fit chi-square formula is used, with the fitted values being the fitted entries to the table arising from the solution to Eq. (23). Zelen (*38*) suggested a rather simple approximate test for interaction in $2 \times 2 \times K$ tables. Halperin et al. (*39*) pointed out that this test was not valid for values of $\psi \neq 1$, although it is good approximation when $\psi \sim 1$ and the n's are not too disproportionate. However, the Bartlett-Norton test is preferred.

Unconditional Inference on ψ for the Combined Case

The analysis of the common ψ in the unconditional space is most easily handled using logit arguments. Woolf (*33*) used logits to find point estimators, interval estimators, and interaction tests for this model. Haldane (*40*), Anscombe (*41*), Gart and Zweifel (*42*), and Cox (*21*), modified and improved somewhat the approximations to the mean and variance of the logit. Gart (*29*) showed the logit point estimator to be efficient and derived two other efficient point and interval estimators for the common ψ. Plackett (*43*), Grizzle, Starmer, and Koch (*44*), and Cox (*21*) extended the arguments of Woolf to more complex situations.

An Application of Combination of Strata Methodology

It has long been known that the incidence of nasopharyngeal carcinoma (NPC) has an elevated incidence among Chinese and Chinese-related

populations. This has led to the speculation that this susceptibility has a genetic basis. Simons et al. (*45*) studied its association with the joint occurrence of HL-A2 and SIN-2 in a case control study among the Chinese population of Singapore. The data and analyses are reproduced in Table 5. Note that the strata here refer to dialect spoken by the individuals involved. The combined analyses (last column) clearly shows an association of this genetic trait with NPC. The analyses of the individual dialect groups shows only the Hokkien and Teochew groups with a significant P value. However, the interaction tests do not show any significant differences among the individual estimators of ψ. For comparison purposes the analyses of the pooled estimator is shown in the penultimate row of Table 5. As the probabilities vary little among the strata, the results differ little from the combined analyses. The authors note that the cases with SIN-2 had poorer survival than those without it. Thus it is not thought that survival bias is present in this study. The authors suggest that the HL-A2-SIN-2 occurrence may confer susceptibility to NPC.

Prospective Studies

Binomial Analyses

In prospective studies the proportion of cases or deaths is directly observed in the two populations, say, $\hat{P}_1 = y_1/N_1$ and $\hat{P}_2 = y_2/N_2$. The unconditional maximum likelihood estimates $\hat{R} = \hat{P}_1/\hat{P}_2$. One can test $R = 1$ by the Fisher-Irwin exact test as $P_1Q_2/P_2Q_1 = 1$ is equivalent to $R = 1$. However, for the binomial model, $Y_.$ is not an appropriate ancillary statistic for R (*46*), and thus conditional inference on R is not possible. Buhrman (*47*) showed that when inverse sampling is used (sampling until $N_1 - y_1$ non-cases and $N_2 - y_2$ non-cases are found in the respective populations), exact conditional confidence limits for R can be derived (*48, 49*). These do not appear to be useful in the usual epidemological problem. Thomas and Gart (*50*) derived "exact" limits for R from the exact limits on ψ assuming all the marginals fixed. These limits have the required confidence coefficient in the conditional space if P_1 and P_2 are such that $N_1P_1 + N_2P_2 \sim y_.$, but they may be below the required value otherwise. Katz et al. and Gart and Thomas (*23*) investigated the Gart-Thomas limits in the unconditional space and found them yield true confidence coéfficients near the nominal values. Katz et al. also suggested basing limits on the logtransform, i.e., let $\ln \hat{R} = \ln \hat{P}_1 - \ln \hat{P}_2$ be approximately normally distributed, and found useful limits on R in a fashion analogous to Woolf's limits on ψ.

The logarithmic transformation may be used in the unconditional space to analyze the combination of R's over several strata. Combined estimators, confidence limits, and "interaction" (equality of the R) tests follow in a manner completely analogous to Woolf's (*33*) results. Radhakrishna (*32*) also investigated the robustness and power of the combined test of a common R.

Analysis of *R* in the Poisson Model

As most diseases have a low incidence and usually prospective studies are concerned with large populations, it is appropriate to approximate the distribution of the numbers of cases by the Poisson distribution. If we also assume an exponential regression of the mean of the Poisson variables on the population and strata effects (*52, 53*), the model yields minimal

Table 5. Analyses of a case-control study of the association between HL-A2/Sin-2 occurrence and nasopharynglol carcinoma in Chinese populations.[a]

	Proportion HL-A2/SIN-2 (%)			Exact analyses (95% limits)	
Dialect	Cases	Controls	$\hat{\psi}_{cml}$	$\neq$	ψ
Cantonese	13/33 (39%)	8/30 (27%)	1.77	0.2115	$0.545 \le \psi \le 6.06$
Hokkien and Teochew	19/54 (35%)	5/43 (12%)	4.07	0.0064	$1.29 \le \psi \le 15.47$
Other	7/23 (30%)	5/16 (31%)	0.964	0.6138	$0.199 \le \psi \le 4.93$
Pooled totals	39/110 (35%)	18/89 (20%)	2.22	0.0103	$1.12 \le \psi \le 4.53$
Combined analyses	Exact interaction test: $p = 0.2584$ Approximate interaction test: $p = 0.2385$		2.17	0.0124	$1.09 \le \psi \le 4.45$

[a]Data of Simons et al. (*45*).

sets of sufficient statistics. Consider the several strata case with Poisson means,

$$E(Y_{1k}) = N_{1k} \exp \{\beta_k + (\alpha/2)\}$$
$$E(Y_{2k}) = N_{2k} \exp \{\beta_k - (\alpha/2)\}$$
$$k = 1, 2, \ldots, K \qquad (28)$$

This model assumes that $R = P_{1k}/P_{2k} = e^{\alpha}$ ($k = 1, 2, \ldots, K$), that is, the R is constant over strata. The likelihood of this model yields $K + 1$ sufficient statistics $y_{1.}$ and $y_{.k}$, $k = 1, 2, \ldots, K$. Conditioning on the $y_{.k}$ yields a product of K independent binomial distributions exactly analogous to the hypergeometric distributions found in the binomial analyses of ψ. The details of the estimation of the common R, exact and approximate tests of $R=1$ [cf. standardized mortality ratio (*54*)], and tests of the model are given in Gart (*55*).

Gart (*55*) extended this model to comparing populations and adjusting for the sexual composition of the two populations. Letting the second subscript denote sex (1 = male, 2 = female), we assume that,

$$E(Y_{11k}) = N_{11k} \exp \{\beta_k + \sigma_k + (\alpha/2)\}$$
$$E(Y_{12k}) = N_{12k} \exp \{\beta_k - \sigma_k + (\alpha/2)\}$$
$$E(Y_{21k}) = N_{21k} \exp \{\beta_k + \sigma_k - (\alpha/2)\}$$
$$E(Y_{22k}) = N_{22k} \exp \{\beta_k - \sigma_k - (\alpha/2)\} \qquad (29)$$

where σ_k is the sex effect on the incidence. This model assumes the population ratios within sexes are constant over age and strata, that is,

$$R_{.1k} = P_{11k}/P_{21k} = e^{\alpha}$$
$$= P_{12k}/P_{22k} = R_{.2k}$$
$$k = 1, 2, \ldots, K \qquad (30)$$

The sex ratios are constant within strata, but may vary over strata [see Gart (*55*) for details of the analyses of this model, particularly for testing $\alpha = 0$, the lack of population differences.]

Analysis of the Ratios of *R* in the Poisson Model

The more interesting model to consider is whether the ratio of population ratios is constant over the sexes

$$\psi_k(\mathbf{P}) = R_{.1k}/R_{.2k}$$
$$= (P_{11k}\, P_{22k})/(P_{21k}\, P_{12k}) = \psi\,(\mathbf{P})$$
$$k = 1, 2, \ldots, K \qquad (31)$$

This can be tested by considering the exponential model:

$$E(Y_{11k}) = N_{11k} \exp \{\beta_k + \alpha_k + \sigma_k + (\Delta/4)\}$$
$$E(Y_{12k}) = N_{12k} \exp \{\beta_k + \alpha_k - \sigma_k - (\Delta/4)\}$$
$$E(Y_{21k}) = N_{21k} \exp \{\beta_k - \alpha_k + \sigma_k - (\Delta/4)\}$$
$$E(Y_{22k}) = N_{22k} \exp \{\beta_k - \alpha_k - \sigma_k + (\Delta/4)\}$$
$$k = 1, 2, \ldots, K \qquad (32)$$

Clearly $\psi_k(\mathbf{P}) = \exp\{\Delta\}$, and tests on Δ are equivalent to tests on $\psi(\mathbf{P})$. This model possesses the minimal set of sufficient statistics, $y_{11.}, y_{..k}, y_{.1k}$, and $Y_{1.k}, K = 1, 2, \ldots, K$. The conditional distribution of $y_{11.}$ given all the other sufficient statistics yields a product of hypergeometric distributions in parallel to the conditional distribution for the common odds ratio [see Eqs. (6) and (22)]. The fixed marginals in the K 2×2 tables here are $y_{1.k}, y_{2.k}, y_{.1k}$, and $y_{.2k}$. The parameter is not ψ but $\psi(p)\,\psi_k(N)$, where

$$\psi_k(N) = \frac{N_{11k}N_{22k}}{N_{12k}N_{21k}} \qquad k = 1, 2, \ldots, K \qquad (33)$$

If the $\psi_k(N) = 1$, for k, the two analyses are exactly equivalent with the total cases playing the role of the n. However if they are not, the test of H_0: $\psi(\mathbf{P}) = 1$ involves a noncentral distribution of each of K hypergeometric variates. For this case, Gart (*55, 56*) has derived the detailed tests of $\psi(\mathbf{P}) = 1$, point estimators, and tests of the model. If all the $\psi_k(N) \sim 1$, the binomial analysis and this analysis will yield quite similar results. If one is comparing disproportionate populations, such as native-born and immigrant populations, the $\psi_k(N)$ may depart considerably for unity and the Poisson analysis is the more appropriate.

Applications of the Poisson Model

We consider two cases in which population based data are used to test hypotheses concerning cancer incidence.

Example 1: Non-melanoma Skin Cancer. Scotto et al. (*57*) studied the incidence of skin cancer other than melanoma among whites in four areas of the U.S.A. Latitude greatly effects the incidence of this disease. It is of interest to compare the incidence in the northernmost population studied, Minneapolis-St. Paul, with the southernmost, Dallas-Fort Worth, These data are presented in Table 6. Clearly Dallas-Fort Worth has higher rates than Minneapolis-St. Paul. For males R the average ratio is $R_{.1.} = 2.73$ ($z = 23.38, p < 0.0001$) and for females, the average ratio is $R_{.2.} = 2.23$ ($z = 15.80, p < 0.0001$). It is also clear that male incidence is higher than the female incidence. However, is the ratio of area incidences higher for males than for females? Or equivalently is the relative risk for males to females higher in the southern population than in the northern population? This is answered by testing whether $\psi(P) = 1$ (or $\Delta = 0$). The asymptotic maximum likelihood estimator of $\psi(P)$ is $\hat{\psi}(P) = 1.19$, which, it should be noted, is not simply the ratio of the $\hat{R}$. The

combined normal deviate test yields $z = +2.44$ ($p \sim 0.0073$). Table 6 also gives the individual estimates and tests by age group. In two of the early age groups and the oldest age group $\psi_k(P)$ is less than one. These are based on such small numbers as not to be significantly different from one. The three ages groups spanning 45-75 yield $\psi_k(P)$ appreciably greater than one and each yields a normal deviate test which is nearly significant. There is not, however, any significant variation among the $\psi_k(P)$ as the goodness of fit test is nowhere near significance. These data clearly show that the North-South incidence ratio is significantly higher for males than for females.

It should be noted that these populations have roughly proportionate distributions by sex. The $\psi_k(N)$'s vary from 1.055 in the youngest age group to 0.910 in the oldest group. Thus we may expect the usual binomial analysis, when applied to the cases (y_{11k} plays the role of x_{1k}; y_{21k}, the role of x_{2k}, $y_{1.k}$, the role of $n_{.1k}$, $y_{2.k}$, the role of n_{2k}, etc.) to yield similar results. In applying the binomial analysis we find $\hat{\psi} = 1.18$ (as opposed to 1.19) and $z = 2.28$ ($p \sim 0.0112$) as opposed to $z = 2.45$ ($p \sim 0.0073$).

Example 2: Fluoridation. A recent report in the U.S. Congressional Record (*58*) attempted to link cancer to the artificial fluoridation of the water supplies. In examining this question, Hoover et al. (*59*) used, among other data, the cancer incidence data of the Second (60) and Third (*61*) National Cancer Surveys. The second survey was done in 1947-48 and the third survey was done in 1969-71. Hoover et al. noted that Denver was not fluoridated in 1947-48, but by 1955, 66% of the area was fluoridated. On the other hand, Birmingham was largely unfluoridated throughout the time period, being only 3.2% fluoridated in 1970. Thus if fluoridation has an effect on cancer incidence we might expect the rates in Denver to increase relative to those in Birmingham over this time period.

As once again we have a comparison of a northern and southern city Hoover et al. (*59*) excluded the skin cancers whose rates, we have just seen, are greatly affected by differences in latitude. As male ratios are more likely to be affected by occupational considerations, we shall consider here only the famel rates (although Hoover et al. considered both). The data are given in Table 7. It is to be noted that the second survey was of one year duration while the third survey was a three-year survey. This difference does not affect the comparisons within surveys or the analyses of the ratio of ratios, but it would affect the direct comparison of surveys. In Table 7 we should note that the cities play the roles the sexes previously played and the surveys play the roles the populations previously played.

Using the binomial analyses, Hoover et al found the $\hat{\psi}$ (Denver to Birmingham) for the second survey to be 1.02, and for the third survey to be 1.07. Since the 95% confidence limits (*17, 37*) for ψ in the second survey (0.91-1.15) entirely covered those for the third survey (1.02-1.13), they took this as "indicating no statistically significant differences."

The populations considered here are not proportionately distributed by survey time; the $\psi_k(N)$ vary somewhat. They range from 0.539 in the youngest age group to 1.274 in the oldest age group. Thus the Poisson analysis is preferred when testing $\psi(P) = 1$.

The results in Table 7 agree, in the main points, with Hoover et al. The relative risks of Denver to Birmingham is found to be 1.02 in the second survey and 1.07 in the third survey. The asymptotic ML estimator of the ratio of ratios is found to be 0.969. The normal deviate test of $\psi(P) = 1$ yield $z = -0.46$

Table 6. Analysis of non-melanoma skin cancer incidences by age and sex in Dallas-Fort Worth and Minneapolis-St. Paul.[a]

	Dallas-Fort Worth		Minneapolis-St. Paul		City Ratios		Ratio of	Test of ψ_k (P) = 1			Normal
Age	Male	Female	Male	Female	$\hat{R}_{.1k}$	$\hat{R}_{.2k}$	$\hat{\psi}_k(P)$	y_{11k}	$\hat{E}(Y_{11k}/\Delta = 0)$	$\hat{V}(Y_{11k}/\Delta = 0)$	z_k
15-24	2/164,065	4/181,343	2/148,099	1/172,675	0.90	3.81	0.24	2	2.69	0.4927	−0.27
25-34	42/147,153	38/146,207	11/122,056	16/123,065	3.17	2.00	1.58	42	39.70	5.0458	0.80
35-44	179/120,195	119/121.374	50/ 95,490	30/ 96,216	2.84	3.14	0.90	179	180.50	15.0569	−0.26
45-54	409/107,558	221/111,353	95/ 87,076	71/ 92,051	3.49	2.57	1.35	409	399.53	30.6131	1.62
55-64	556/ 73,342	259/ 83,004	150/ 62,268	102/ 72,159	3.15	2.21	1.43	556	540.28	43.2614	2.31
65-74	480/ 38,868	310/ 55,932	165/ 37,391	130/ 54,722	2.80	2.33	1.20	480	470.51	51.8550	1.25
75-84	261/ 15,739	226/ 29,007	165/ 19,012	133/ 32,195	1.91	1.89	1.01	261	260.40	45.7775	0.01
84+	47/ 3,360	65/ 7,538	32/ 4,081	40/ 8,328	1.78	1.80	0.99	47	47.07	10,7619	−0.02
					$\hat{R}_{.1.}$ = 2.73	$\hat{R}_{.2.}$ = 2.24	$\hat{\psi}(P)$ = 1.19	1976	1940.69	204.8644	$z = 2.44$

Goodness-of-fit chi-square = 6.316 (7 d.f.)

[a]Data of Scotto et al. (*57*).

Table 7. Comparison of the temporal changes of the total cancer incidence in Denver and Birmingham in females by age.[a]

Age	Second Survey		Third Survey		City ratios				Test of $\psi_k(P) = 1$		Normal deviate
	Denver	Birmingham	Denver	Birmingham	2nd survey	3rd survey	Ratio of ratios				
k	y_{11k}/N_{11k}	y_{12k}/N_{12k}	y_{21k}/N_{21k}	y_{22k}/N_{22k}	$\hat{R}_{1.k}$	$\hat{R}_{2.k}$	$\psi(P)$	y_{11k}	$\hat{E}(Y_{11k}/\Delta=0)$	$\hat{V}(Y_{11k}/\Delta=0)$	z_k
0-14	6/59,583	6/44,466	49/167,152	12/65,960	0.75	1.61	0.46	6	7.74	2.1836	−0.84
15-24	17/37,935	6/28,669	77/109,387	31/45,503	2.14	1.03	2.07	17	14.05	4.3281	1.18
25-34	39/44.580	33/30,500	153/ 83,897	68/34,465	0.81	0.92	0.87	39	40.73	12.9060	−0.34
35-44	80/37,011	70/25,887	396/ 70,569	196/33,498	0.80	0.96	0.83	80	85.27	28.8148	−0.89
45-54	172/28,907	107/18,993	757/ 64,817	413/33,326	1.06	0.94	1.12	172	165.93	53.6524	0.76
55-64	206/22,772	111/13,002	912/ 46,017	462/27,671	1.06	1.19	0.89	206	212.56	57.0846	−0.80
65-74	188/15,487	85/ 7,844	902/ 31.664	522/19,381	1.12	1.06	1.06	188	185.12	50.4732	0.34
74+	135/ 7,929	55/ 3,337	966/ 23,429	431/12,246	1.03	1.17	0.88	135	139.26	33.0855	−0.65
					$\hat{R}_{1..}$ = 1.02	$\hat{R}_{2..}$ = 1.07	$\hat{\psi}(P)$ = 0.969	843	850.65	242.5282	z = −0.46

Goodness-of-fit Chi-Square = 6.484 (7d.f.)

[a]Data of Dorn and Cutler (*60*), Cutler and Young (*61*), and Hoover et al (*59*).

($p \sim 0.323$). The goodness of fit test yields a chi-square of 6.484, with 7 degrees of freedom, so there is no indication of that the $\psi_k(p)$'s vary significantly over the age groups. Thus this test also concludes that there is no indication of a significant increase in the cancer rates among females in Denver as compared to Birmingham in the period following fluoridation in Denver. This test also yields a similar non-significant finding for the male comparison.

I am grateful to Joseph Scotto, John L. Young, and Robert N. Hoover for making available to me the detailed data used in the last two examples. I am grateful to Alroy M. Smith for computer programming and Sue Tiffany for typing of the manuscript.

REFERENCES

1. McKinlay, S. M. The design and analysis of the observational study — a review. *J. Am. Stat. Assoc.* 70; 503 (1975).
2. Fisher, R. A. The logic of inductive inference. J. Roy. Statist. Soc. A98: 39 (1935).
3. Cornfield, J. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. J. Nat. Cancer Inst. 11: 1269 (1951).
4. Cornfield, J. A statistical problem arising from retrospective studies. Proc. 3rd *Berkeley Symp.* Vol. 4, J. Neyman, Ed., Univ. Calif. Press, Berkeley, 1956, pp. 135-148.
5. Doll, R., and Hill, A. B. Smoking and carcinoma of the lung. Preliminary report, Brit. Med. J. 2: 739 (1950).
6. Miettinen, O. S. Estimability and estimation in case-referrent studies. Am. J. Epid. 103: 226 (1976).
7. Miettinen, O. S. Reply to letter to editor. Am. J. Epid. 105: 498 (1977)
8. Halperin, M. Letter to the editor. Am. J. Epid. 105: 496 (1977).
9. Bartlett, M. S. Contingency table interactions. J. Roy. Statist. Soc. B 2: 248 (1935).
10. Norton, H. W. Calculation of chi-square for complex contingency tables. J. Am. Statist. Assoc. 40: 251 (1945).
11. Armitage, P. The use of the cross-ratio in aetiological surveys. in: Perspectives in Probability and Statistics, J. Gani, Ed., Academic Press, New York-London, 1975, pp. 349-355.
12. Lehmann, E. L. Testing Statistical Hypotheses. Wiley, New York, 1959, Chapt. 4.
13. Fisher, R. A. Confidence limits for a cross-product ratio. Austral. J. Statist. 4: 41 (1962).
14. Cox, D. R. The regression analysis of binary sequences (with discussion). J. Roy. Statist. Soc. B 20: 215 (1958).
15. Birch, M. W. The detection of partial association. I. The 2 × 2 case. J. Roy. Statis. Soc. B26: 313 (1964).
16. Thomas, D. G. Exact confidence limits for an odds ratio in a 2 × 2 table. Appl. Statist. 20: 105 (1971).
17. Thomas, D. G. Exact asymptotic methods for the combination of 2 × 2 tables. Comp. Biomed. Res. 8: 423 (1975).
18. Stevens, W. L. Mean and variance of an entry in a contingency table. Biometrika 38: 468 (1951).
19. Gart, J. J. The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. Rev. Int. Statis. Inst. 39: 148 (1971); Corrigendum 40: 221 (1972).
20. Gart, J. J., and Thomas, D. G. Numerical results on approximate confidence limits for the odds ratio. J. Roy. Statist. Soc. B34; 441 (1972).
21. Cox, D. R. The Analysis of Binary Data. Methuem, London, 1970.
22. Gart, J. J. Approximate confidence limits for the relative risk. J. Roy. Statist. Soc. B24: 454 (1962).
23. Gart, J. J., and Thomas, D. G. Evaluation of confidence limits for the odds ratio and the difference and ratio of proportions. In preparation.
24. Bross, I. D. J. Misclassification in 2 × 2 tables. Biometrics 10: 478 (1954).
25. Goldberg, J. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. J. Am. Statist. Assoc. 70: 561 (1975).
26. Copeland, K. T., Checkoway, H., McMichael, A. J., and Holbrook, R. H. Bias due to misclassification in the estimator of the relative risk. Am. J. Epid., 105: 488 (1977).
27. Cox, D. R. Two further applications of a model for binary regression. Biometrika 45: 562 (1958).
28. Siegel, D. G., and Greenhouse, S. W. Validity in estimating the relative risk in case-control studies. J. Chron. Dis. 26: 219 (1973).

29. Gart, J. J. On the combination of relative risks. Biometrics 18, 601 (1962).
30. Rogentine, G. N., Jr., Yankee, R. A., Gart, J. J., Nam, J., and Trapani, R. J. HL-A Antigens and disease. Acute lymphocytic leukemia. J. Clin. Invest. 51: 2420 (1972).
31. Rogentine, G. N., Trapani, R. J., Yankee, R. A., and Henderson, E. S. HL-A Antigens and acute lymphocytic leukemia: the nature of the HL-A Association. Tissue Antigens 3: 470 (1973).
32. Radhakrishna, S. Combination of results from several 2×2 contingency tables. Biometrics 21: 86 (1965).
33. Woolf, B. On estimating the relationship between blood group and disease. Ann. Hum. Genet. 19: 251 (1955).
34. Cochran, W. G. Some methods strengthening the common χ^2 tests. Biometrics 10: 417 (1954).
35. Cox, D. R. A simple example of a comparison involving quantal data. Biometrika 53: 215 (1966).
36. Mantel, N., and Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. J. Nat. Cancer Inst. 22: 719 (1959).
37. Gart, J. J. Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals. Biometrika 57: 471 (1970).
38. Zelen, M. The analysis of several 2×2 contingency tables. Biometrika 58: 129 (1971).
39. Halperin, M., Ware, J. H., Byar, D. P., Mantel, N., Brown, C. C., Koziol, J., Gail, M., and Green, S. B. Testing for interaction in a $I \times J \times K$ contingency table. Biometrika 64: 271 (1977).
40. Haldane, J. B. S. The estimation and significance of the logarithm of a ratio of frequencies. Ann. Hum. Genet., 20: 309 (1955).
41. Anscombe, F. J. On estimating binomial response relations. Biometrika 43: 461 (1956).
42. Gart, J. J., and Zweifel, J. R. On the bias of the logit and its variance with application to quantal bioassay. Biometrika 54: 181 (1967).
43. Plackett, R. L. A note on interactions in contingency tables. J. Roy. Statist. Soc. B 24: 162 (1962).
44. Grizzle, J. E., Starmer, C. F., and Koch, G. G. Analysis of categorical data by linear models. Biometrics 24: 489 (1969).
45. Simons, M. J., Wee, G. B., Goh, E. H., Chan, S. H., Shanmugaratnam, K., Day, N. E., and de-Thé, G. Immunogenetic aspects of nasopharyngeal carcinoma. IV. Increased risk in Chinese of nasopharyngeal carcinoma associated with a Chinese-related HLA profile (A2, Singapore 2). J. Nat. Cancer Inst. 57: 977 (1976).
46. Barndorff-Nielson, O. Nonformation. Biometrika 63: 567 (1976).
47. Buhrman, J. M. Tests and confidence intervals for the difference and ratio of two probabilities. Biometrika 64: 160 (1977).
48. Barton, D. E. W. Query 21. Comparison of sample sizes in inverse binomial sampling. Technometrics 9: 337 (1967).
49. Kudo, A., and Tarumi, T. 2×2 table emerging out of different chance mechanisms. Comm. Stat. Theor. Meth., A7: 977 (1978).
50. Thomas, D. G., and Gart, J. J. A table of exact confidence limits for difference and ratios of two proportions and their odds ratios. J. Am. Stat. Assoc. 72: 73 (1977); Corrigendum, ibid. 73: 233 (1978).
51. Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. Obtaining confidence limits for the risk ratio in cohort studies. Biometrics 33: 469 (1978).
52. Armitage, P. The chi-square test for heterogeneity of proportions after adjustment for stratification. J. Roy. Statist. Soc. B28: 150 (1966); Addendum, ibid. 29: 197 (1967).
53. Cox, D. R. Notes on some aspects of regression analyses (with discussion). J. Roy. Statist. Soc. A131: 265, 315 (1968).
54. Armitage, P. Statistical Methods in Medical Research. Wiley, New York, 1971.
55. Gart, J. J. The analysis of ratios and cross-product ratios of Poisson variates with application to incidence rates. Comm. Stat. Theor. Meth. A7: 917 (1978).
56. Gart, J. J. The Poisson distribution: The theory and application of some conditional tests. In: Statistical Distributions in Scientific Work Vol. 2, G. P. Patil, S. Kotz, and J. K. Ord, Eds., D. Reidel, Boston, 1975, pp. 125-140.
57. Scotto, J., Kopf, A. W., and Urbach, F. Non-melanoma skin cancer among Caucasians in four areas of the United States. Cancer 34: 1333 (1974).
58. Congressional Record: House H7172-H7176, July 21, 1975.
59. Hoover, R. N., McKay, F. W., and Fraumeni, J. F. Fluoridated drinking water and the occurrence of cancer. J. Nat. Cancer Inst. 57: 757 (1976).
60. Dorn, H. F., and Cutler, S. J. Morbidity from Cancer in the United States (Public Health Monograph 56), D.C., Washington, U.S. GPO, 1959.
61. Cutler, S. J., and Young, J. L., Jr. Third National Cancer Survey. Incidence Data. U.S. GPO, Washington, D.C., 1975.